

JGR Atmospheres

RESEARCH ARTICLE

10.1029/2019JD031232

Key Points:

- The performance for day 1 O₃ forecasts driven by RAP is generally improved than that driven by GFS
- The reduction in day 1 O₃ forecast errors is mainly associated with the boundary layer height
- The cloud assimilation improves cloud forecasts for the first few hours and reduces errors in 1-hr O₃ forecast under cloudy conditions

Supporting Information:

- Supporting Information S1

Correspondence to:

A. Hodzic,
alma@ucar.edu

Citation:

Ryu, Y.-H., Hodzic, A., Descombes, G., Hu, M., & Barré, J. (2019). Toward a Better Regional Ozone Forecast Over CONUS Using Rapid Data Assimilation of Clouds and Meteorology in WRF-Chem. *Journal of Geophysical Research: Atmospheres*, 124, 13,576–13,592. <https://doi.org/10.1029/2019JD031232>

Received 25 JUN 2019

Accepted 10 NOV 2019

Accepted article online 14 NOV 2019

Published online 8 DEC 2019

Author Contributions:

Conceptualization: Alma Hodzic, Gael Descombes, Jérôme Barré
Formal analysis: Young-Hee Ryu
Investigation: Young-Hee Ryu, Alma Hodzic
Methodology: Young-Hee Ryu, Gael Descombes
Resources: Ming Hu
Supervision: Alma Hodzic, Gael Descombes
Validation: Young-Hee Ryu
Visualization: Young-Hee Ryu
Writing - original draft: Young-Hee Ryu
Writing - review & editing: Alma Hodzic, Gael Descombes, Jérôme Barré

©2019. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Toward a Better Regional Ozone Forecast Over CONUS Using Rapid Data Assimilation of Clouds and Meteorology in WRF-Chem

Young-Hee Ryu¹ , Alma Hodzic¹ , Gael Descombes^{1,2}, Ming Hu³, and Jérôme Barré⁴

¹National Center for Atmospheric Research, Boulder, CO, USA, ²Now at INERIS, Verneuil-en-Halatte, France, ³NOAA/OAR/Earth System Research Laboratory, and Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA, ⁴European Centre for Medium-Range Weather Forecasts, Reading, UK

Abstract Accuracy of cloud predictions in numerical weather models can considerably impact ozone (O₃) forecast skill. This study assesses the benefits in surface O₃ predictions of using the Rapid Refresh (RAP) forecasting system that assimilates clouds as well as conventional meteorological variables at hourly time scales. We evaluate and compare the WRF-Chem simulations driven by RAP and the Global Forecast System (GFS) forecasts over the Contiguous United States (CONUS) for 2016 summer. The day 1 forecasts of surface O₃ and temperature driven by RAP are in better agreements with observations. Reductions of 5 ppb in O₃ mean bias error and 2.4 ppb in O₃ root-mean-square-error are obtained on average over CONUS with RAP compared to those with GFS. The WRF-Chem simulation driven by GFS shows a higher probability of capturing O₃ exceedances but exhibits more frequent false alarms, resulting from its tendency to overpredict O₃. The O₃ concentrations are found to respond mainly to the changes in boundary layer height that directly affects the mixing of O₃ and its precursors. The RAP data assimilation shows improvements in the cloud forecast skill during the initial forecast hours, which reduces O₃ forecast errors at the initial forecast hours especially under cloudy-sky conditions. Sensitivity simulations utilizing satellite clouds show that the WRF-Chem simulation with RAP produces too thick low-level clouds, which leads to O₃ underprediction in the boundary layer.

1. Introduction

Ground-level ozone (O₃) is a secondary pollutant formed by chemical reactions that involve nitrogen oxides (NO_x) and volatile organic compounds (VOCs) in the atmosphere under the presence of ultraviolet (UV) radiation. Because of its adverse effects on human health and on ecological systems, O₃ is subject to regulations outlined in the National Ambient Air Quality Standards. The current air quality standard for ground-level O₃ in the United States is 70 ppb for maximum daily 8-hr average O₃ (MDA8 O₃). To forecast and regulate O₃, photochemical air quality models have been used to simulate the physical and chemical processes that affect O₃ as well as other air pollutants. Despite considerable efforts made in the last two decades to improve O₃ forecast skill, accurate O₃ prediction still remains challenging (Ryan, 2016). A recent evaluation of an ensemble of operational regional air quality models over the Contiguous United States (CONUS; Im et al., 2015) showed that models tend to overestimate the summertime surface MDA8 O₃ levels below 50 ppb and underestimate levels above 70 ppb that are of greatest concern for air quality forecast.

One of the challenges lies in the uncertainties in meteorology, and several previous studies reported large O₃ model errors associated with uncertainties in meteorological variables (2–10 ppb for MDA8 O₃ in summertime over the United States; Zhang et al., 2007; Gilliam et al., 2015; Foley et al., 2015). For example, Gilliam et al. (2015) examined the impact of inherent meteorology uncertainty on O₃ prediction using ensemble perturbation techniques and several different meteorological models. They showed that MDA8 O₃ can have a standard deviation of 10 ppb, which represents 15–20% of the mean ensemble value. They highlighted the large ensemble variability associated with the planetary boundary layer (PBL) height and shortwave radiation (closely related to clouds) whose spreads are 25% of the mean ensemble value and 20% of the maximum incoming radiation, respectively.

Table 1
WRF-Chem Model Configuration and Physics and Chemistry Options

Process	Method
Horizontal grid size	12 × 12 km ² (420 × 273 points)
Vertical levels	50
Pressure top	15 hPa
Lateral boundary conditions	GFS forecast (0.5° resolutions)
Time step	60 s
Microphysics	Thompson (Thompson et al., 2008)
Shortwave and longwave radiation	RRTMG (Iacono et al., 2008)
Cumulus parameterization	Grell and Freitas (2014)
Boundary layer	MYNN (Nakanishi & Niino, 2006, 2009)
Land surface	RUC (Benjamin et al., 2004)
Gas phase chemistry	MOZART-4 (Knote et al., 2014)
Photolysis	Updated TUV (v5.3) with RRTMG COD

Among many meteorological variables influencing O₃ concentration, UV radiation plays a key role in O₃ formation as it directly drives the photochemistry. Hence, clouds that significantly modulate UV radiation in the atmosphere have been recognized as one of the most important meteorological factors. Recently, Ryu et al. (2018) utilized satellite cloud retrievals to estimate O₃ uncertainties associated with the uncertainties in cloud predictions using the Weather Research and Forecasting with chemistry (WRF-Chem) model. They showed that the summertime averaged MDA8 O₃ can have errors of 1–5 ppb caused by incorrect clouds in the model simulations, which represent up to about 40% of the total MDA8 O₃ bias under cloudy conditions. However, this method that consisted of replacing the radiative forcing of modeled clouds with satellite clouds is not applicable to O₃ forecasts.

From the perspective of numerical weather prediction (NWP), there have been various attempts to assimilate or initialize clouds in NWP models,

leading generally to improved cloud predictions as reviewed by van der Veen (2013). The improvement in cloud forecasts has been generally found to last only several hours, typically for 1–7 hr. A recent study by Chen et al. (2016) demonstrated that the forecast with variational assimilation of satellite cloud water/ice paths shows a significant improvement in precipitation for the first 7 hr. However, to the best of our knowledge, the effects of cloud assimilation in O₃ predictions have not been fully quantified so far.

In this context, we use the National Oceanic and Atmospheric Administration (NOAA) Rapid Refresh (RAP) hourly analysis, which use nonvariational cloud data assimilation, as meteorological inputs to drive WRF-Chem forecasts every hour and to estimate its effects on summertime O₃ forecasts over CONUS. The RAP is an hourly updated assimilation and model forecast system based on the WRF model, which has been used for operational short-range weather forecasts since 2012 (Benjamin et al., 2016). The objectives of the present study are threefold: (1) to evaluate O₃ forecast skill utilizing RAP and compare the performance with that using the Global Forecast System (GFS) forecasts as meteorological inputs, (2) to assess the benefits of O₃ forecasts from RAP hourly updated meteorological and cloud fields, and (3) to identify the relative contributions of meteorological factors to O₃ forecasts.

2. Methods

2.1. Experiment Design

WRF-Chem version 3.9.1.1 is used in this study. The model configuration is identical to that used in RAP version 3 (hereafter RAPv3) except for the model domain and the grid resolutions. Our model domain covers CONUS and is embedded within the larger RAPv3 domain over North America. The WRF-Chem model uses the horizontal grid of 12 km and 50 vertical levels with a model top pressure of 15 hPa, whereas RAPv3 uses 13.545 km horizontal grid with 51 vertical layers extending up to 10 hPa. The simulation period is 6 June 2016 through 31 August 2016, but the June simulations are used for the spin-up and discarded in the present study. This ample spin-up period is considered to provide proper initial conditions for some species that are not obtained from the MOZART global simulation. The model setup for gas-phase chemistry, chemical initial and boundary conditions, and anthropogenic/biogenic emissions is similar to that used in Ryu et al. (2018) except for the omission of aerosols and the calculation of cloud optical depth (COD) in the photolysis rates. In the present study, only gas-phase chemistry (with MOZART-4 mechanism, see Knote et al., 2014, and references therein) is considered to reduce computational costs. Unlike Ryu et al. (2018) in which COD is computed inside the tropospheric ultraviolet and visible (TUV) photolysis scheme using the formulation based on Chang et al. (1987) and Briegleb (1992) in WRF-Chem, the present study uses CODs from the RRTMG shortwave radiation scheme for consistency between physics and chemistry. The model configuration with key physics and chemistry options used in the present study is summarized in Table 1.

As in RAPv3, the meteorological lateral boundary conditions are provided by 3 hr and beyond GFS forecasts with analysis time of 00 and 12 UTC. In other words, the lateral boundary conditions are updated at 03 and 15 UTC. Because RAPv3 has hourly assimilation frequency, the WRF-Chem forecast that uses RAPv3 as

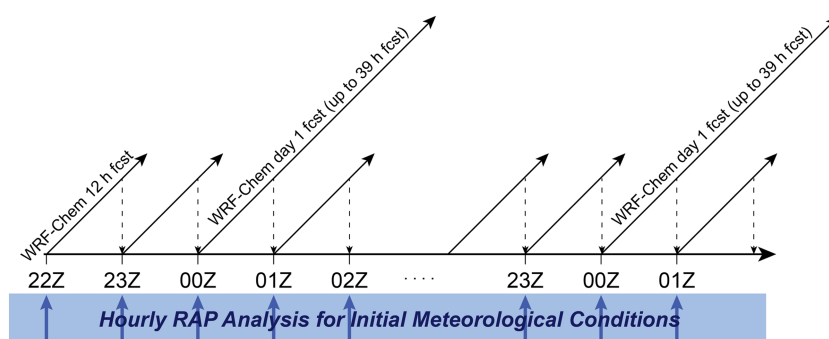


Figure 1. A schematic diagram of hourly updated WRF-Chem-RAP cycling. The initial concentrations of chemical species of each forecast cycle are obtained from the 1-hr forecast of previous cycle. Day 1 forecast is initialized at 00 UTC and produced up to 39 hr.

meteorological initial conditions (hereafter, WRF-Chem-RAP) is launched every hour using the hourly updated RAPv3 analysis. We used RAPv3 analysis with its native grids on sigma levels. Figure 1 illustrates a schematic of how WRF-Chem-RAP forecasts are conducted. For each WRF-Chem-RAP run, 12-hr forecasts are produced except for the one that is initialized at 00 UTC. For 00 UTC initialized WRF-Chem-RAP, 39 h forecasts are produced because this run is used for the day 1 forecast. According to Ryan (2016), in most locations the air quality forecast must be issued in the late morning or early afternoon for public warnings and more importantly to encourage private and public entities to curb emissions. This timing constraint leads to the necessity for the air quality forecast to be a 12- to 36-hr forecast. Considering the fact that more meteorological observations are available at 00 and 12 UTC from rawinsonde observations than other times of day and considering the constraints of the public warnings, we start WRF-Chem-RAP forecasts at every 00 UTC for day 1 forecast. Chemical species are cycled between hourly launched WRF-Chem-RAP runs; that is, their initial concentrations are obtained from the 1-hr forecast of the previous run that has been initialized 1 hr before.

For data assimilation of cloud and precipitation hydrometeors in RAPv3, which is nonvariational data assimilation, the surface observations of cloud ceiling and visibility and satellite-based retrievals of cloud top pressure and temperature are used to modify background hydrometeor fields. In addition, the water vapor and temperature fields are adjusted by conserving virtual potential temperature to retain the cloud or clear information in the forecast. Finally, three-dimensional radar data are used to clear and build hydrometeors such as rain, snow, and graupel (Benjamin et al., 2016).

In addition to clouds, in RAPv3, many variables are assimilated using observational data from surface, rawinsonde, wind profilers, aircraft, and buoy/ship measurements. It is noteworthy that near-surface observations from METARs and to a lesser extent mesonet are used to assimilate surface variables such as pressure, temperature, dew point temperature, and horizontal winds. The RAP soil-atmosphere coupled data assimilation particularly provides a significant constraint on the land surface variables such as soil temperature and soil moisture, which are adjusted based on the increments of temperature and relative humidity at the lowest model level, respectively. A more detailed description of RAPv3 data assimilation is given by Benjamin et al. (2016).

For the control simulation, we use the same model configuration as in WRF-Chem-RAP, but the model is driven by the GFS forecasts (0.5° on pressure levels) for both initial and lateral boundary meteorological conditions (hereafter, WRF-Chem-GFS). The GFS forecast is often used for air quality forecast as meteorological initial and boundary conditions (Bei et al., 2008; Saide et al., 2016; Zhou et al., 2017). For consistency with RAPv3 in which 3-hr GFS forecasts valid at 03 and 15 UTC are used as its background field (Benjamin et al., 2016), the WRF-Chem-GFS starts at 03 and 15 UTC each day using 3-hr GFS forecasts. By doing so, we intend to quantify the effect of RAPv3 data assimilation on the WRF-Chem O_3 predictions. The WRF-Chem-RAP forecast is expected to differ from the WRF-Chem-GFS forecast not only in terms of clouds but also in terms of other meteorological variables. Therefore, the main meteorological state variables that largely affect levels of atmospheric pollutants, which are temperature, dew point temperature, horizontal

Table 2
Control (WRF-Chem-GFS) and Sensitivity (WRF-Chem-RAP, WRF-Chem-GOES, WRF-Chem-HRRR) Simulations

Simulation Characteristics	WRF-Chem-GFS	WRF-Chem-RAP	WRF-Chem-GOES	WRF-Chem-HRRR
Initial conditions	GFS forecast	RAP forecast	RAP forecast	HRRR forecast
Cloud cover	WRF-generated	WRF-generated	GOES satellite	WRF-generated
Forecast cycle	03 and 15 UTC	Every hour	Every hour	Every hour
Initialization time for day 1 forecast	03 UTC	00 UTC	00 UTC	00 UTC

wind components, and boundary layer height, are evaluated against observations, and the performance of the WRF-Chem simulations driven by RAP and GFS forecasts are compared.

A set of sensitivity simulations in which GOES satellite cloud retrievals are used to constrain clouds for photochemistry (hereafter WRF-Chem-GOES) is conducted to further examine its effects on O_3 prediction. The WRF-Chem-GOES uses the same meteorology as WRF-Chem-RAP. The same method used in Ryu et al. (2018) is adopted, which replaces WRF-generated clouds with GOES satellite clouds only in the calculation of actinic flux and irradiation. In this case, only photolysis rates and photosynthetically active radiation for emissions of biogenic VOCs are influenced by GOES satellite clouds, and the meteorology is not influenced by the satellite clouds.

We conduct another set of sensitivity simulations, which use the High Resolution Rapid Refresh (HRRR) as meteorological inputs, to examine the effects of high-resolution weather forecasts on surface O_3 forecast. The HRRR explicitly resolves clouds and convection at 3 km and employs more frequent radar assimilation at 15-min intervals. Because of extensive computational demand for running the WRF-Chem model at 3-km horizontal resolutions, we opt to run the WRF-Chem model at 12-km resolutions but use the 3-km HRRR forecasts as initial meteorological conditions (hereafter, WRF-Chem-HRRR). The model configuration in terms of physics and chemistry and hourly updated WRF-Chem simulations are identical to those used for WRF-Chem-RAP. A shorter period of 2–16 August 2016 is considered for the sensitivity simulation using HRRR, and the initial conditions for chemical species are taken from the 1-hr forecast of WRF-Chem-RAP valid at 00 UTC 2 August 2016. The description of the experiments performed in the present study is summarized in Table 2.

2.2. Observations

2.2.1. Meteorological Data

To evaluate the meteorology performance of WRF-Chem-GFS and WRF-Chem-RAP, we use the hourly surface observations of temperature, dew point temperature, wind speed, and wind direction that are obtained from NOAA's Meteorological Development Laboratory (previously, Techniques Development Laboratory) and that are archived at the National Center for Atmospheric Research. In the present study, data from 2,056 surface sites are used for the model evaluation over CONUS, which include synoptic observations, meteorological aviation reports, automated weather observing systems, and automated surface observing systems.

2.2.2. Planetary Boundary Layer Height Data

The PBL height and temperature observations at the Atmospheric Radiation Measurement Program Southern Great Plains (ARM SGP) site are used to validate the modeled PBL height and temperature in WRF-Chem-GFS and WRF-Chem-RAP. The ARM observations using radiosonde profiles are available at 4 times of the day (at 00, 06, 12, and 18 UTC) and provide four types of PBL height estimates. In the present study, we use the PBL height estimates based on the Liu and Liang (2010) method, which is considered as a reference in Molod et al. (2015). Note that we have used only the ARM data set at the SGP site to evaluate the model performance because to the authors' best knowledge this is the only data set that provides routine and quality-controlled PBL height estimates over CONUS.

2.2.3. EPA Surface O_3 Data

The U.S. Environmental Protection Agency (EPA) hourly O_3 measurements at 1,299 stations are used to evaluate the modeled O_3 . Given strong spatial differences in O_3 precursors and chemistry over CONUS, seven subregions are considered following Eder et al. (2009), that is, Northeast (NE), Upper Midwest (UM), Lower Midwest (LM), Southeast (SE), Rocky Mountain (RM), Pacific Coast (PC), and California (CA; see Figure S1 in supporting information).

2.2.4. GOES Satellite Cloud Data

As in Ryu et al. (2018), satellite cloud retrievals of cloud bottom height, cloud top height, and COD from GOES 13 (GOES-East) and GOES 15 (GOES-West) are used to evaluate the model performance and also to conduct the WRF-Chem-GOES simulation in this study. This data set is a subset of the hourly 8-km cloud retrievals performed using the Satellite Cloud and Radiation Property Retrieval System (Minnis et al., 2011). The 8-km cloud retrievals are regridded to 12-km WRF-Chem grids using the nearest-neighbor interpolation as done in Ryu et al. (2018). Uncertainties associated with the satellite products are described in Ryu et al. (2017).

2.3. Evaluation Methods

2.3.1. Cloud Contingency

The modeled cloudiness is evaluated against the GOES satellite cloud retrievals. A standard 2×2 contingency table consisting of four categories, which are hit (a), false alarm (b), miss (c), and correct negative (d), is used for the model evaluation (Table S1, supporting information). For both the satellite data and model simulations, the grid cell is labeled as cloudy (clear) when hourly COD at a grid is greater than or equal to (smaller than) 0.3. The threshold value of 0.3 used in this study is slightly higher than the lowest detection limit of satellite retrieved COD over land of 0.25. In the present study, two additional metrics are used to evaluate cloud forecast performance: (1) Accuracy = $a + d$ and (2) Probability of detection (POD) = $a/(a + c)$.

2.3.2. Categorical Statistics for O₃ Evaluation

To obtain MDA8 O₃ at an EPA observation site, we take 8-hr running averages of O₃ starting from local midnight to 11 p.m. local time at the site (so, there are 23 values of 8-hr average of O₃). Then, the maximum value among the 23 eight-hour averages are selected for MDA8 O₃ at the site. The metrics used for categorical O₃ forecast evaluation are based on Eder et al. (2006), but the notations used in the present study are different from those used in Eder et al. (2006) to make them consistent with the notations for cloud evaluation. A data point in which both observation and forecast exceed the MDA8 O₃ standard (70 ppb) is denoted by hit (a). The false alarm (b) denotes that a forecast exceeds the standard but the observation does not. The miss (c) is the case in which a forecast does not exceed the standard but the observation does. A data point in which neither forecast nor observation exceeds the standard is denoted by correct negative (d). In addition to POD, two metrics are used to evaluate O₃ forecast skill, which are the false alarm ratio ($FAR = b/(a + b)$) and the critical success index ($CSI = a/(a + b + c)$). The CSI represents the hit probability, given that the event was forecast and/or observed, and it is particularly suited to rare and extreme events (Lawson et al., 2018).

To evaluate O₃ forecast skill in the presence of significantly thick clouds, we defined cloudy-sky conditions as follows: COD at an EPA observation site is greater than 20 (to consider only thick clouds) at least for 4 hr during the 8-hr time window that is used for computation of MDA8 O₃. For clear sky, a stricter condition is applied: no clouds or very thin clouds (COD < 0.3) are observed during the period of 13–01 UTC that corresponds to 08–20 EST or 05–17 PST.

3. Model Evaluation

3.1. Surface Meteorological Variables

The conventional surface meteorological observations for 2-m temperature, 2-m dew point temperature, 10-m wind speed, and 10-m wind direction are used to evaluate the meteorology performance (Figure 2). The day 1 forecasts valid at daytime hours (10–17 local standard time) are considered here. Figure 2 shows that WRF-Chem-GFS underestimates temperature with a mean bias error of -0.76 °C while WRF-Chem-RAP performs significantly better with a much smaller mean bias error of 0.04 °C. The root-mean-square error (RMSE) and correlation coefficient are also improved in WRF-Chem-RAP compared to WRF-Chem-GFS. The WRF-Chem-RAP particularly shows an improved performance in capturing high temperatures exceeding ~ 32 °C. The dew point temperature is also better predicted in WRF-Chem-RAP than in WRF-Chem-GFS in terms of mean bias error, RMSE, and correlation coefficient. The same analyses are performed for the seven subregions (Table S2, supporting information). The RMSEs are reduced over all the subregions in WRF-Chem-RAP compared to those in WRF-Chem-GFS. The reduction in bias errors are especially noticeable over the northeast and southeast. Only California exhibits a worse performance in terms of bias errors for 2 m and dew point temperatures in WRF-Chem-RAP than in WRF-Chem-GFS.

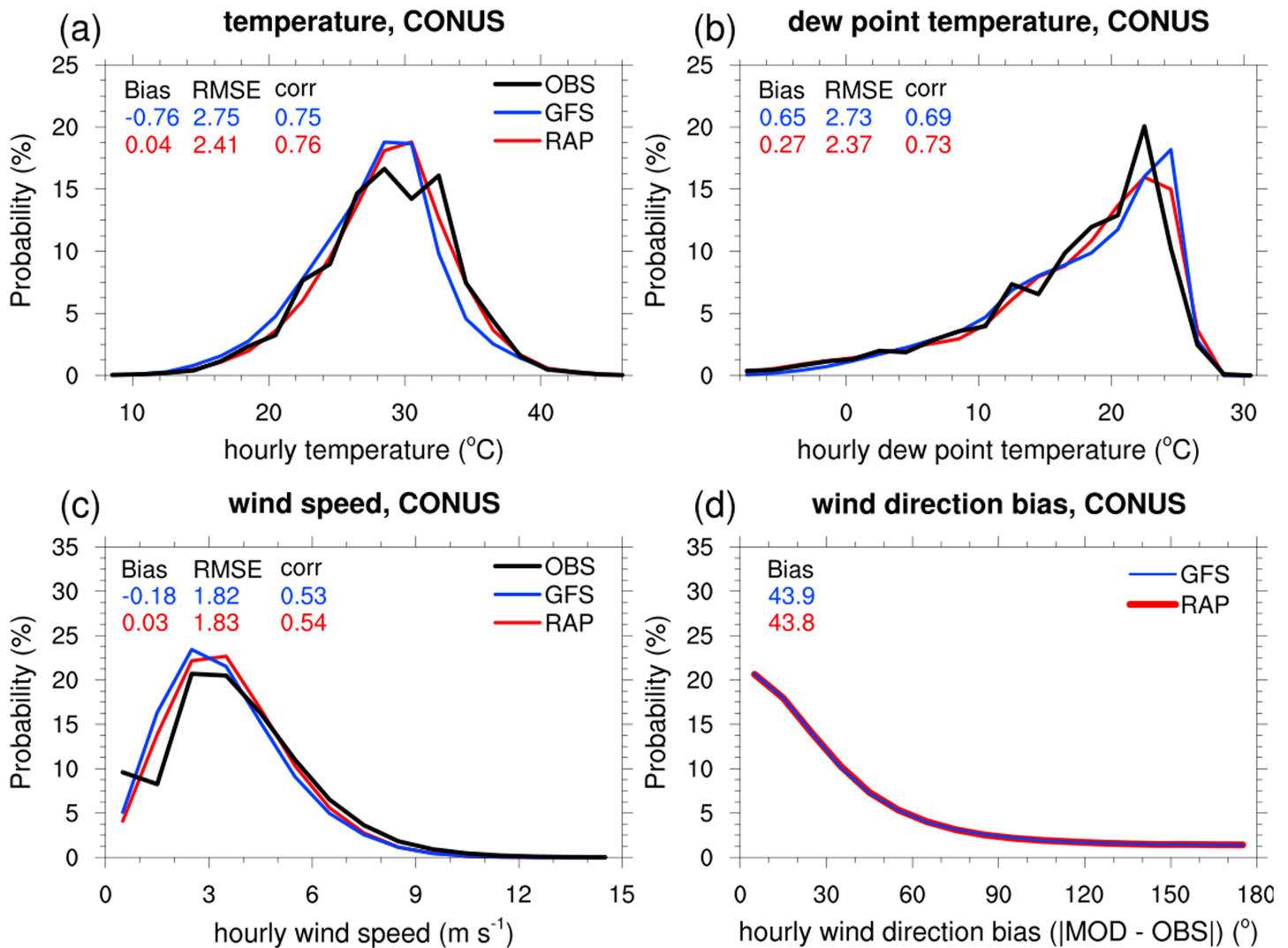


Figure 2. Probability density functions of near-surface meteorological variables from observations and day 1 forecasts of WRF-Chem-GFS and WRF-Chem-RAP during 1 July to 31 August 2016 over CONUS. (a) 2-m temperature, (b) 2-m dew point temperature, (c) 10-m wind speed, and (d) 10-m wind direction bias (absolute difference between observations and forecasts). Note that the wind direction bias in WRF-Chem-RAP (red line) is very similar to that in WRF-Chem-GFS (blue line; the two are almost completely overlapped). The description of surface meteorological observations is given in section 2.2.1.

The wind speed and wind direction show minor improvements in WRF-Chem-RAP as compared to WRF-Chem-GFS. The wind speed probability density function (PDF) indicates that WRF-Chem-GFS tends to slightly underestimate wind speed with a mean bias error of -0.18 m/s, which is better captured in WRF-Chem-RAP. The wind direction bias (defined as the absolute difference in wind direction between the model and observations) shows negligible differences between the two simulations. Overall, the model evaluation shows that WRF-Chem-RAP performs better in capturing the 2016 summer conditions over CONUS than WRF-Chem-GFS especially in terms of near-surface temperatures.

3.2. Planetary Boundary Layer Height at the SGP Site

Figure 3 compares the observations and day 1 forecasts of 2-m temperature and PBL height at 18 UTC (= 12 CST) from WRF-Chem-GFS and WRF-Chem-RAP. Note that due to a small number of data samples (once a day at 18 UTC) the observations and model forecasts for the time period of 6 June 2016 to 31 August 2016 are used and compared. The WRF-Chem-RAP is in a better agreement with observations for 2-m temperature with smaller mean bias error and RMSE, which is consistent with the results shown above

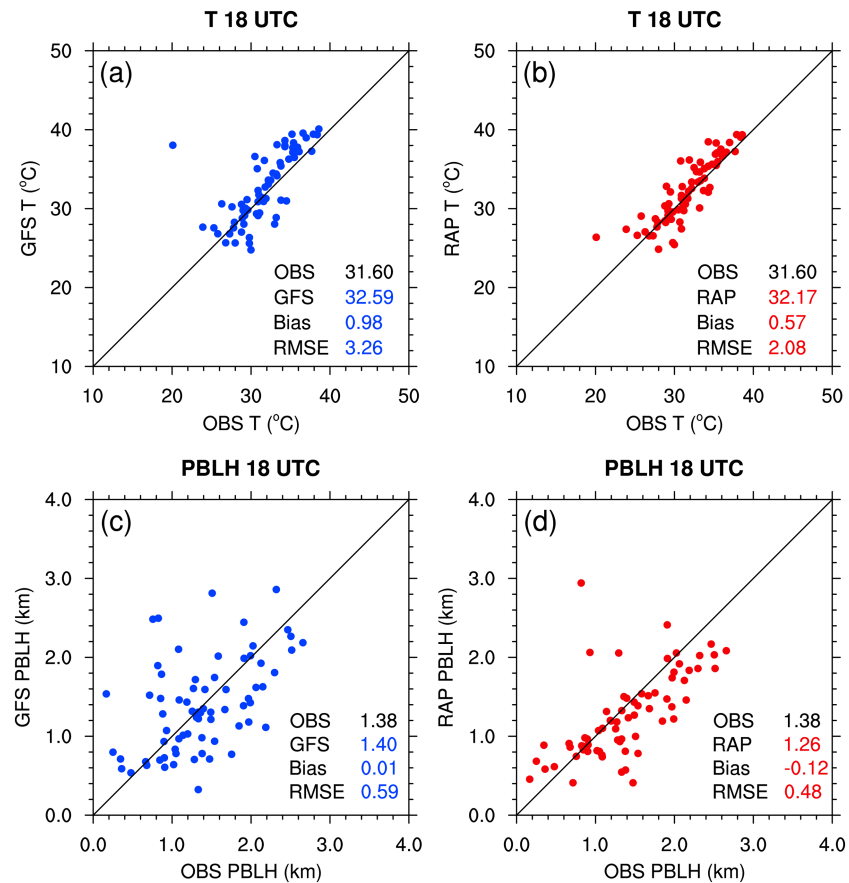


Figure 3. Scatter plots of (a) day 1 WRF-Chem-GFS forecast of 2-m temperature at 18 UTC (= 12 CST) and (c) day 1 WRF-Chem-GFS forecast of planetary boundary layer height (PBLH) at 18 UTC at the SGP site. Panels (b) and (d) are the same as panels (a) and (c), respectively, but for WRF-Chem-RAP forecasts. The mean values from observations, WRF-Chem-GFS, and WRF-Chem-RAP are denoted by OBS, GFS, and RAP, respectively. The description of PBLH observations is given in section 2.2.2.

(Figure 2, section 3.1). In terms of PBL, the WRF-Chem-RAP predicts 10% lower PBL height compared to WRF-Chem-GFS, resulting in a slight (8%) underestimation of the observed values but at the same time in a reduction in RMSE. The scatter plots suggest that a better performance in predicting 2-m temperatures leads to a better performance in predicting PBL heights. It should be noted, however, that the differences in 2-m temperature and PBL height between WRF-Chem-GFS and WRF-Chem-RAP at the SGP site are rather small and of opposite sign compared to most other regions (see Figure 8; the star marker indicates the location of the SGP site).

3.3. Clouds

3.3.1. Day 1 Cloud Forecast

The daytime hourly cloudiness of day 1 forecasts (valid at 16–23 UTC, 13–23 hr later after the initialization) from WRF-Chem-GFS and WRF-Chem-RAP is evaluated through a 2×2 contingency table (Table 3). The comparison shows that there are only small differences between the two simulations for all four categories (hit, false alarm, miss, and correct negative). The WRF-Chem-RAP shows a slightly better performance in hit and in miss, which are 23.5% and 20.5%, respectively, as compared to those in WRF-Chem-GFS (22.8% for hit and 21.7% for miss). However, the false alarm is slightly higher and correct negative is slightly lower in WRF-Chem-RAP than in WRF-Chem-GFS. The higher hit and false alarm rates in WRF-Chem-RAP indicate that WRF-Chem-RAP produces ~5% more frequent cloudiness than WRF-Chem-GFS. The probability of detection is also slightly higher in WRF-Chem-RAP (53.4%) than in WRF-Chem-GFS (51.2%).

Table 3
Cloud Contingency Table for WRF-Chem-GFS and WRF-Chem-RAP
Day 1 Forecasts

		GOES satellite	
		Cloudy	Clear
Model	Cloudy	Hit	False alarm
	GFS	22.8%	8.53%
	RAP	23.5%	9.44%
Clear	Miss	Correct negative	
	GFS	21.7%	47.0%
	RAP	20.5%	46.6%

Note. Clouds over land are only considered, and hourly daytime CODs (16–23 UTC) are averaged. The values in the table are normalized by total number of data, so that the unit is percentage (%).

A more quantitative analysis on COD and cloud top heights is shown in Figure 4, illustrating the bivariate PDF for GOES satellite retrievals and WRF-Chem-GFS and WRF-Chem-RAP day 1 cloud forecasts. The satellite cloud retrievals show that the most abundant clouds appear at low-level (cloud top height below ~3 km) with CODs of 1–10 and their peak value of ~3. The low-level clouds with thin to moderate CODs correspond to cumulus and/or stratocumulus clouds (Hahn et al., 2001). The CODs of stratocumulus widely vary ranging from less than 1 to more than 20 and locally can exceed 50 (Wood, 2012). The second distinctive distribution centered at COD of ~3 at high levels is the characteristic of thin cirrus clouds with top heights of 8–12 km. These results are consistent with the findings obtained from GOES 2013 cloud retrievals in Ryu et al. (2018). Neither of the WRF-Chem simulations captures the characteristics of cloud distribution found in satellite retrievals. In fact, the two simulations show rather similar distributions. Both simulations predict a larger

number of thinner clouds than observed (CODs <1 instead of 1–10) although the top height of high-level and low-level clouds is well captured. This result could be caused by the coarse model resolution (12 km), which is not able to explicitly resolve small-scale clouds in the horizontal direction. For thick clouds with CODs of 10–50, both simulations show more frequent low-level clouds compared to satellite retrieved clouds. This distribution could be a result of a shift in low-level clouds whose CODs are overestimated by the Thompson microphysics scheme or/and the Grell-Freitas cumulus scheme. A further study is required to identify causes of this overestimation.

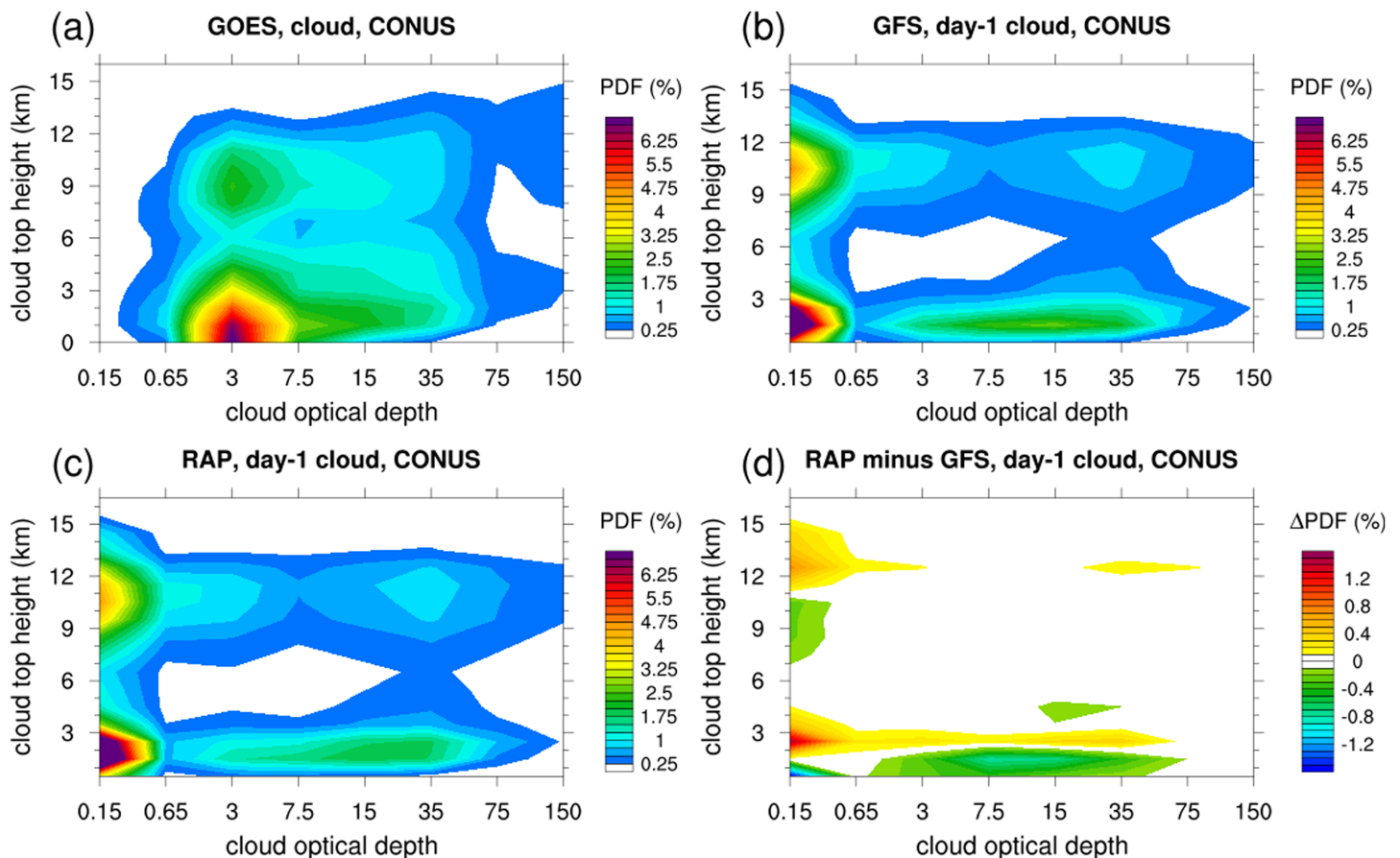


Figure 4. Bivariate probability density function (PDF) of cloud top height (in km, above ground level) and cloud optical depth from (a) GOES satellite retrievals, (b) WRF-Chem-GFS day 1 forecasts, (c) WRF-Chem-RAP day 1 forecasts, and (d) differential PDF between WRF-Chem-RAP and WRF-Chem-GFS day 1 forecasts (WRF-Chem-RAP minus WRF-Chem-GFS). Daytime clouds valid at 16–23 UTC over land are only considered.

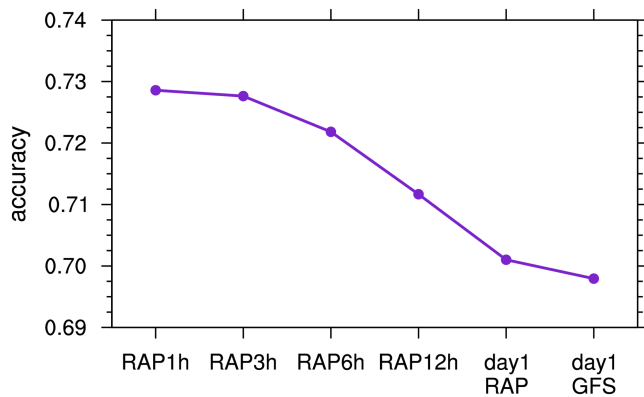


Figure 5. Daytime cloudiness performance in terms of accuracy (= hit rate + correct negative) with respect to 1-, 3-, 6-, and 12-hr WRF-Chem-RAP forecasts, and day 1 WRF-Chem-RAP and WRF-Chem-GFS forecasts.

The differential PDF between two model simulations (Figure 4d) indicates that the clouds in WRF-Chem-RAP have somewhat higher cloud top heights for both low-level (~3 km) and high-level (12–15 km) clouds than those in WRF-Chem-GFS. In addition, more low-level clouds with CODs of 1–75 are found in WRF-Chem-GFS than in WRF-Chem-RAP. Nonetheless, the overall distributions are similar to each other, which implies that the day 1 forecasts for clouds are mostly governed by model dynamics and physics rather than initial conditions.

3.3.2. Hourly WRF-Chem-RAP Cloud Forecasts

Figure 5 shows the accuracy rates for 1, 3, 6, and 12 hr and day 1 WRF-Chem-RAP and WRF-Chem-GFS cloud forecasts for daytime (16–23 UTC) cloudiness. As expected, the accuracy rate is highest for 1-hr forecasts, and decreases with lead time. The accuracy rate decreases from 72.9% for 1 hr, and 70.1% for day 1 WRF-Chem-RAP forecasts. As seen in Table 3, the day 1 forecast performance is slightly better in WRF-Chem-RAP than in WRF-Chem-GFS. It is interesting to note that the 3-hr forecast

performance is similar to the 1-hr forecast performance, implying that the better initial meteorology including better initial clouds can be retained for ~3 hr. This result is consistent with the findings from previous studies (e.g., Bayler et al., 2000; Jones et al., 2013; Martinet et al., 2014; Yucel et al., 2002; Yucel et al., 2003). A similar result was reported by Bytheway and Kummerow (2015) for HRRR whose performance in convective precipitation is best at forecast hour 3. The bivariate PDFs for 1-, 3-, 6-, and 12-hr WRF-Chem-RAP forecasts (Figure S2, supporting information) indicate that the 1-hr forecasts capture the satellite-retrieved cloud PDF slightly better, but this improvement does not last long (~3 hr).

3.4. Surface O₃

The day 1 MDA8 O₃ forecast for the period of 1 July to 31 August 2016 is evaluated against the EPA O₃ surface measurements (Table 4). Over CONUS, the day 1 O₃ forecast of WRF-Chem-GFS is overestimated with a mean bias error of 5.1 ppb. Over all subregions, the day 1 MDA8 O₃ forecasts in WRF-Chem-RAP are lower than those in WRF-Chem-GFS, leading to a considerable improvement in model performance in general. The reductions in mean bias error by 5 ppb and in RMSE by 2.4 ppb in WRF-Chem-RAP compared to WRF-Chem-GFS over CONUS are particularly remarkable, which correspond to 10% and 4.9% reductions with respect to mean MDA8 O₃ in WRF-Chem-GFS, respectively. The results are however contrasting between the eastern and western CONUS regions. The improvements are particularly significant over the eastern United States (northeast, upper midwest, southeast, and lower midwest regions). The maximum reduction in mean bias error by 6.1 ppb is found in the upper midwest, while the largest reduction in RMSE by 5.4 ppb is found in the northeast region. Consistently, the normalized mean bias and normalized mean error are also improved over the same regions, as well as the correlation coefficients except for lower midwest.

Table 4

Performance Statistics of MDA8 O₃ Under All Sky Conditions for 1 July to 31 August 2016 Over CONUS and Subregions

Regions	Mean bias error (ppb)		Normalized mean bias (%)		Root-mean-square error (ppb)		Normalized mean error (%)		Correlation coefficient	
	GFS	RAP	GFS	RAP	GFS	RAP	GFS	RAP	GFS	RAP
CONUS	5.1	−0.09	14.8	2.9	12.6	10.2	25.2	19.7	0.66	0.67
Northeast	11.3	5.8	24.3	12.8	15.2	9.8	27.2	17.1	0.69	0.75
Upper midwest	9.0	2.9	21.7	7.3	12.5	7.8	24.9	15.2	0.71	0.72
Southeast	10.9	6.4	30.1	18.0	13.2	9.0	31.3	21.1	0.75	0.76
Lower midwest	8.0	4.3	25.9	15.6	12.2	10.1	31.5	25.4	0.70	0.65
Rocky Mountain	−8.2	−11.0	−15.2	−20.8	11.0	13.1	18.5	22.3	0.52	0.48
California	−1.3	−9.4	2.9	−13.9	11.5	13.2	20.1	20.9	0.58	0.59
Pacific Coast	−1.2	−3.7	−2.2	−9.6	8.4	7.8	18.8	17.8	0.70	0.73

Note. Only data samples in which MDA8 O₃ is observed in the daytime (with a start time of the 8-hr window ranging from 8 a.m. to 13 p.m. local time) are used for the analysis. For mean bias error, normalized mean bias, root-mean-square error, and normalized mean error, a lower value in absolute magnitude means a better performance. For correlation coefficient, a higher value means a better performance.

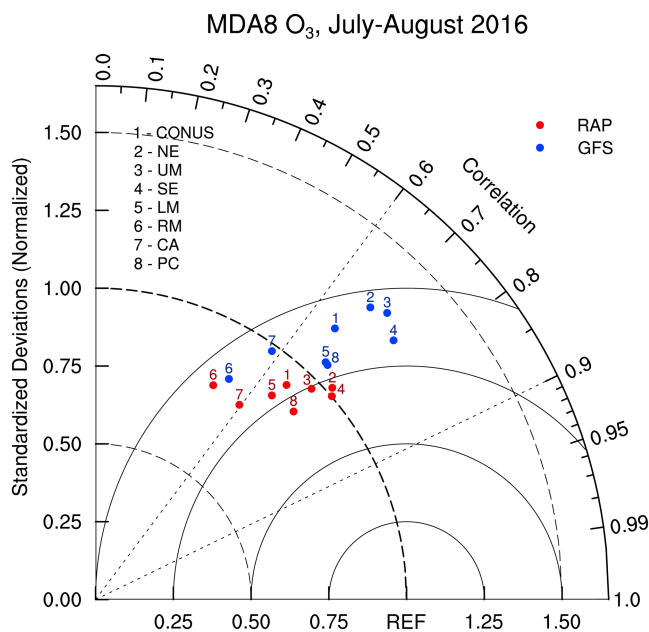


Figure 6. Taylor diagram of day 1 MDA8 O₃ forecasts by WRF-Chem-GFS (blue) and WRF-Chem-RAP (red). The numbers indicate the CONUS or subregions (abbreviations as defined in section 2.2.3 and Table 4).

distances from the reference point denoted by REF) are seen in WRF-Chem-RAP compared to WRF-Chem-GFS over CONUS and subregions except for Rocky Mountain. In particular, northeast, upper midwest, and southeast regions exhibit significantly better improvements when RAPv3 meteorology is used. Therefore, it is reasonable to conclude that the use of RAPv3 meteorology can considerably improve the O_3 forecast skill in terms of RMSE over most regions of CONUS (except for Rocky Mountain and California) for this particular model configuration.

Table 5 shows categorical statistics evaluating O₃ exceedance of the EPA standard (>70 ppb) for MDA8 O₃. Note that over the Pacific Coast no data sample that exceeds the standard is found in WRF-Chem-RAP (so, *a* and *b* explained in section 2.3.2 are zero) for the period of 1 July to 31 August 2016, and hence no categorical statistics are given. Overall, the WRF-Chem-RAP shows a lower probability of detection, but it also shows a lower false alarm ratio than WRF-Chem-GFS. This result is due to the fact that the use of RAPv3 meteorol-

ogy reduces the model overprediction of surface O_3 in most regions as discussed above (see Table 4). It should be noted that WRF-Chem-RAP shows higher critical success index in some subregions than in WRF-Chem-GFS even though the corresponding probability of detections are lower, for example, in northeast, upper midwest, southeast, and lower midwest (those are the regions where the model mean bias error is substantially reduced with RAPv3 meteorology, Table 4).

4. Discussion

4.1. Benefits of Hourly Updated Clouds in WRF-Chem-RAP

To assess the effects of hourly updated clouds and to separate them from hourly updated meteorology (variables other than clouds), the hourly O_3 forecast performance under clear- and cloudy-sky conditions is compared (Figure 7). The underlying hypothesis is that data assimilation for conventional meteorology is applied regardless of cloudiness and that the cloud data assimilation is applied mostly in the presence of clouds or in the likelihood of clouds. This approach provides only an approximate but a reasonable estimate of the effects of cloud data assimilation on O_3 forecast.

Table 5
Categorical Statistics, Which Are Computed Based on the Observed Exceedances, Nonexceedances Versus Forecast Exceedance, Nonexceedances for MDA8 O₃

Regions	Probability of detection (POD, %)		False ALARM RATIO (FAR, %)		Critical SUCCESS INDEX (CSI, %)	
	GFS	RAP	GFS	RAP	GFS	RAP
CONUS	37.8	12.8	81.5	76.1	14.2	9.1
Northeast	83.2	74.8	89.2	77.9	10.6	20.6
Upper midwest	81.5	41.7	92.7	75.9	7.2	18.0
Southeast	88.9	44.4	96.6	91.5	3.4	7.6
Lower midwest	45.5	13.6	94.2	87.0	5.4	7.1
Rocky Mountain	0.9	0.4	90.5	75.0	0.8	0.4
California	32.7	3.2	26.3	11.7	29.2	3.2
Pacific Coast	-	-	-	-	-	-

Note. The standard of exceedance is 70 ppb for MDA8 O₃.

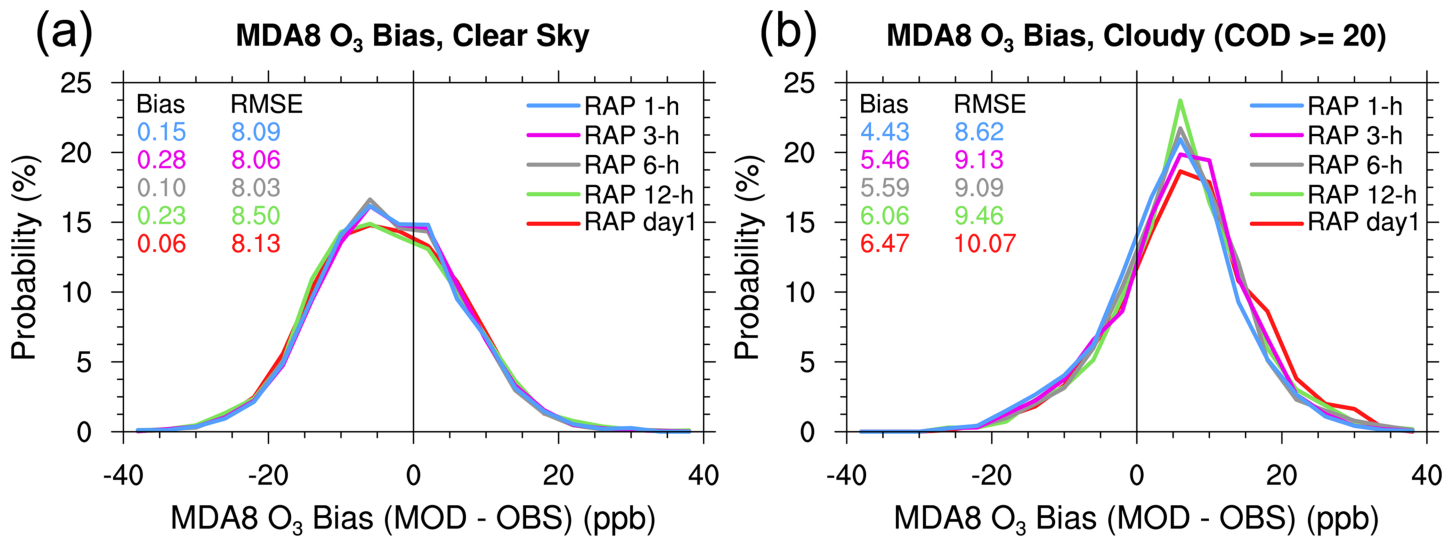


Figure 7. Probability density functions of MDA8 O₃ bias (forecast value minus observation value) for WRF-Chem-RAP 1-, 3-, 6-, and 12-hr forecasts (a) under clear-sky conditions and (b) cloudy-sky conditions over CONUS. Note that California is excluded in this figure because there are only two cloudy-sky samples.

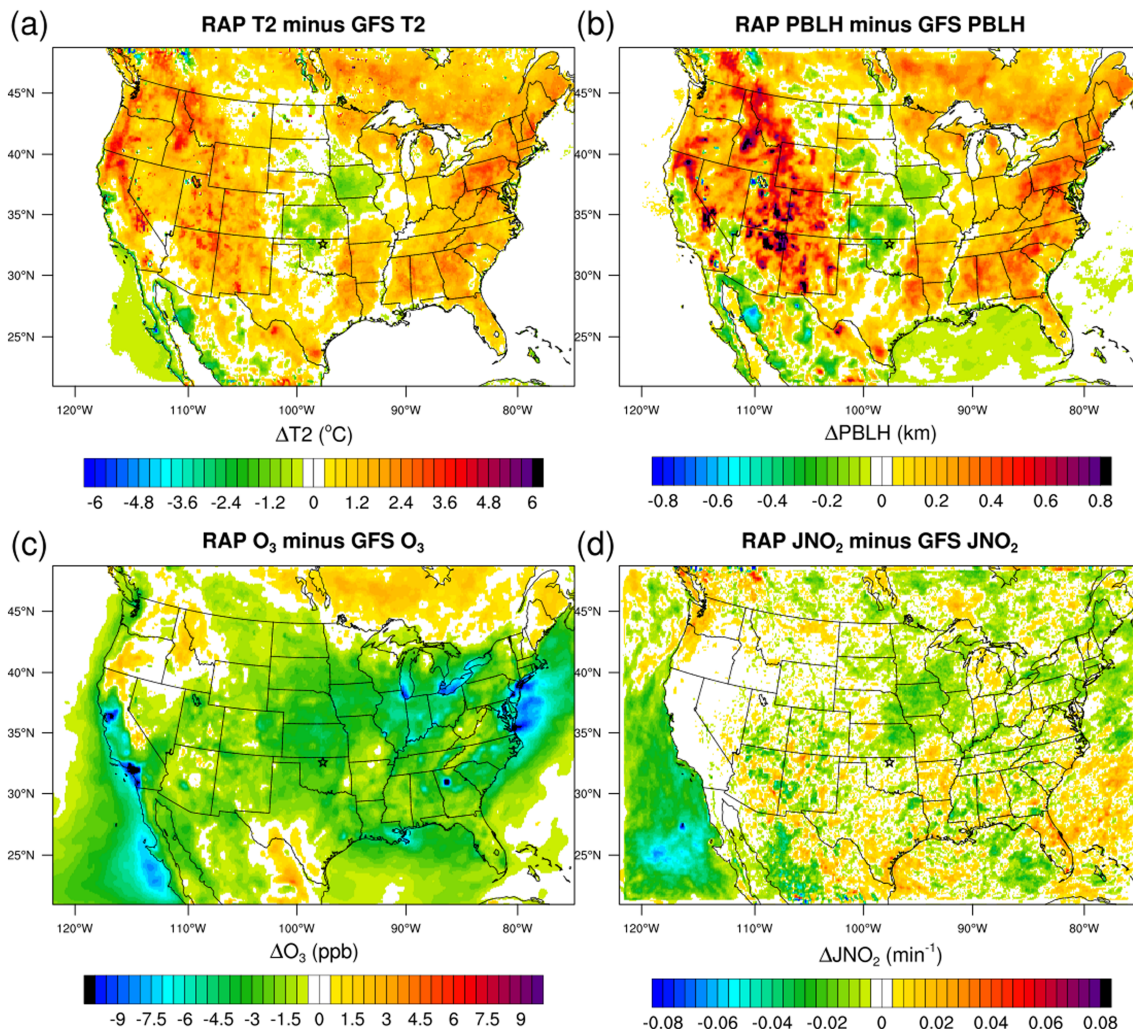


Figure 8. Differential maps of daytime-average (10–17 local standard time) day 1 forecasts of (a) 2-m temperature, (b) PBL height, (c) O₃, and (d) NO₂ photolysis rates (JNO₂) between WRF-Chem-RAP and WRF-Chem-GFS (WRF-Chem-RAP minus WRF-Chem-GFS).

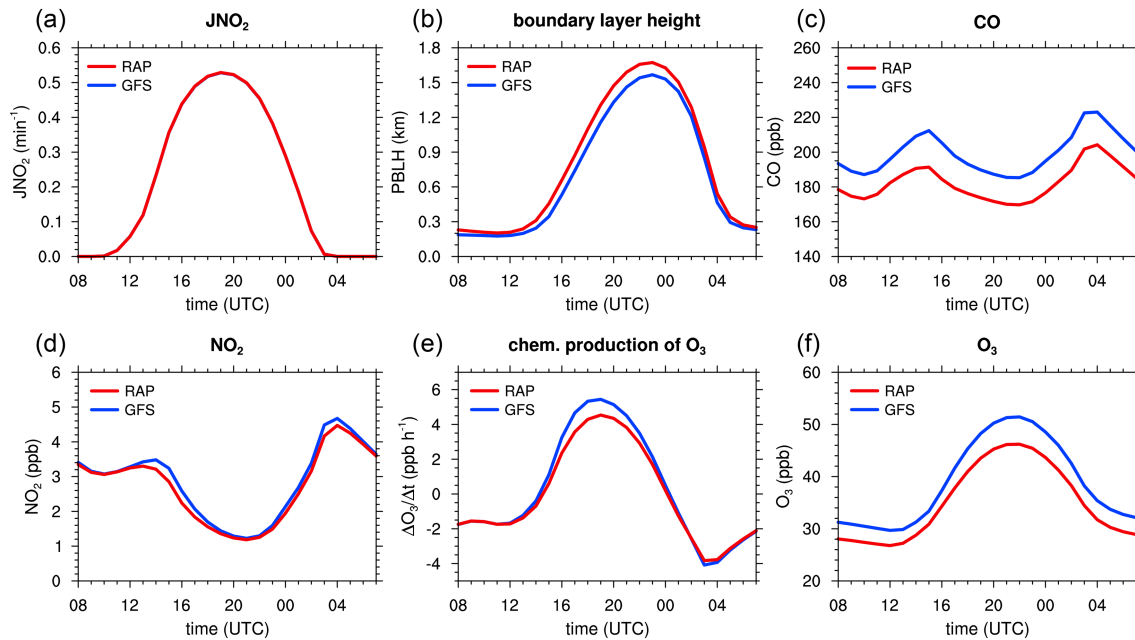


Figure 9. Diurnal variation of clear-sky average of (a) JNO_2 , (b) PBL height, (c) CO, (d) NO_2 , (e) chemical production of O_3 , and (f) O_3 at the lowest model level in WRF-Chem-GFS and WRF-Chem-RAP day 1 forecast.

The exact quantification cannot be achieved here given that RAPv3 forecast products without cloud data assimilation are not available, and thus the effects of cloud data assimilation cannot be separated from those of conventional meteorology data assimilation.

Figure 7 shows that the performance degradation with forecast length in WRF-Chem-RAP is more apparent under cloudy-sky conditions than under clear-sky conditions. For example, the RMSE of MDA8 O_3 under cloudy-sky conditions progressively increases from 8.6 ppb in 1-hr forecast to 9.5 ppb in 12-hr forecast, and up to 10.1 ppb for day 1 forecast. However, the errors under clear-sky conditions do not show such a tendency with forecast length. In fact, the performance of 6-hr forecast is the best in terms of mean bias error and RMSE, although the differences among forecasts are small under clear-sky conditions. These results suggest that hourly updated meteorology has a greater potential to reduce errors under cloudy-sky conditions than under clear-sky conditions.

In addition, the decrease by 1.5 ppb in RMSE between the day 1 MDA8 O_3 forecast and 1-hr forecast under cloudy-sky conditions implies that a benefit of refreshing meteorology every hour is to reduce errors in initial concentrations of surface O_3 . Under clear-sky conditions, the RMSE of day 1 MDA8 O_3 forecast is 8.1 ppb and is comparable to that of 1-hr forecast. One can estimate that the effects of hourly updated RAPv3 meteorology on reducing O_3 RMSE is ~ 0.5 ppb under clear-sky conditions at most, which is the difference between the RMSEs of 6- and 12-hr forecasts. This result can be roughly regarded as the base effect of hourly updated meteorology. The corresponding effect under cloudy-sky conditions is 1.5 ppb at most. Therefore, it can be concluded that the benefit of hourly updated cloud initial fields, which is estimated from the difference in RMSE reduction between clear and cloudy skies, is ~ 1 ppb (1.5 ppb minus 0.5 ppb).

4.2. Effects of Boundary Layer

The most distinctive difference in meteorology between WRF-Chem-RAP and WRF-Chem-GFS is found in 2-m temperature as seen in Figure 2. Figure 8a shows the differential map of daytime-averaged 2-m temperature between the two simulations, and 2-m temperature is generally higher in most regions over CONUS in WRF-Chem-RAP than in WRF-Chem-GFS. This is due to the higher soil temperatures that are adjusted in RAPv3 data assimilation based on near-surface temperature increments (not shown). The higher soil temperatures lead to higher sensible heat fluxes (not shown) and thus higher PBL height in WRF-Chem-RAP than in WRF-Chem-GFS (Figure 8b).

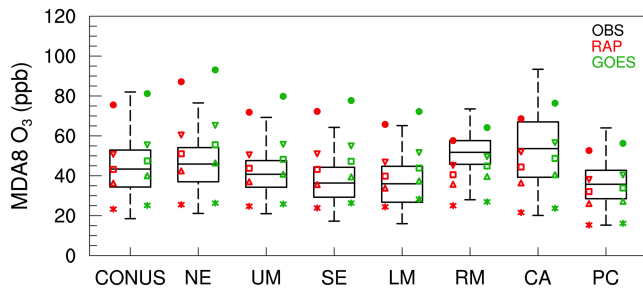


Figure 10. Boxplot of MDA8 O_3 over CONUS and subregions during 1 July to 31 August 2016 in observations (box and whisker in black), WRF-Chem-RAP, and WRF-Chem-GOES. The lower and upper bounds of box in black indicate the 25th and 75th percentiles, respectively, and the center line of box indicates the median of observed MDA8 O_3 . The whiskers indicate the 1st and 99th percentiles of observed MDA8 O_3 . The asterisk, up-pointing triangle, square, down-pointing triangle, and circle indicate 1st, 25th, 50th, 75th, and 99th percentiles of the simulations, respectively.

In Figure 8c, surface O_3 in WRF-Chem-RAP is lower than that in WRF-Chem-GFS over most of CONUS. There could be several reasons for the lower surface O_3 in WRF-Chem-RAP, but the primary reason is the higher PBL height in WRF-Chem-RAP than in WRF-Chem-GFS (Figure 9). It should be noted that both simulations use the MYNN boundary layer parameterization (Table 1). To make the comparison simple, only clear-sky conditions are considered so that NO_2 photolysis rates (JNO_2) exhibit almost identical diurnal variations between two simulations (Figure 9a). The diurnal variation in PBL height shows higher peak values in WRF-Chem-RAP by up to 150 m than in WRF-Chem-GFS. A higher PBL height leads to lower pollutant concentrations in WRF-Chem-RAP due to enhanced dilution within a deeper boundary layer. This effect is illustrated by the diurnal variation in CO (Figure 9c) which clearly shows 15–20 ppb lower concentrations in WRF-Chem-RAP than in WRF-Chem-GFS. CO can be used as an indicator of atmospheric mixing because of its low reactivity and hence relatively long lifetime (~ 2 months) compared to NO_x . The NO_2 concentration is also lower in WRF-Chem-

RAP than in WRF-Chem-GFS (Figure 9d). The lower concentrations of O_3 precursors in the deeper boundary layer lead to the lower chemical production of surface O_3 in WRF-Chem-RAP than in WRF-Chem-GFS (Figure 9e). Together with the dilution of O_3 itself, O_3 concentration in WRF-Chem-RAP is considerably lower by ~ 3 ppb in the nighttime and by ~ 5 ppb in the daytime (Figure 9f). The same analyses are performed for the subregions, and the conclusions are consistent across the subregions (see Figures S4–S10, supporting information).

It should be noted that the negative relationship between PBL height and surface O_3 does not hold for all regions over CONUS. For example, the state of Iowa exhibits the lower PBL height and lower O_3 in WRF-Chem-RAP than in WRF-Chem-GFS (Figures 8b and 8c). This result is in part caused by the lower JNO_2 , which reduces chemical production of O_3 in the daytime (Figure 8d). Optically thicker clouds are found in WRF-Chem-RAP over the state of Iowa and nearby states (Figure S11, supporting information). In addition, the higher PBL height over Iowa in WRF-Chem-GFS (negative differential PBL height over Iowa) could enhance the mixing down of O_3 from the residual layer when boundary layer grows during the daytime, and thus leading to higher O_3 in WRF-Chem-GFS than WRF-Chem-RAP (not shown). It should be noted that temperature may not be a major/direct contributor to the differential O_3 between the two simulations. High temperature acts to increase O_3 concentration. Even though temperature is higher in WRF-Chem-RAP, O_3 is lower in WRF-Chem-RAP than in WRF-Chem-GFS. These results are consistent with previous studies (e.g., Wilczak et al., 2009).

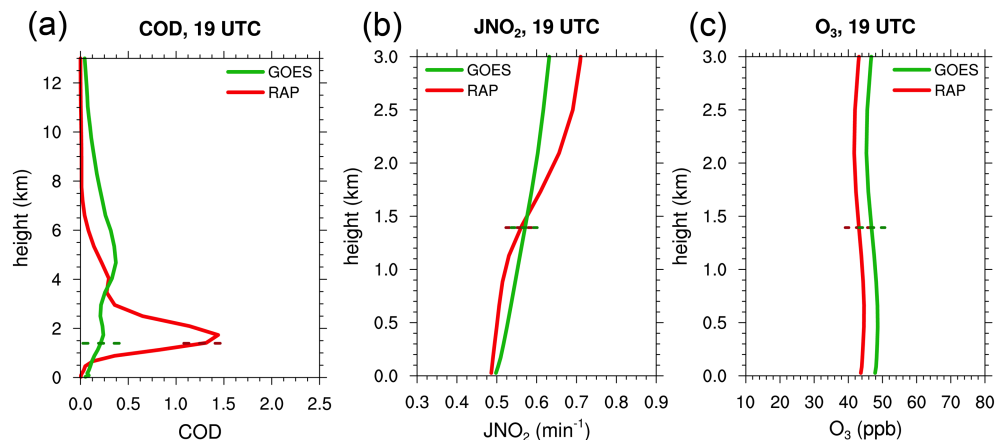


Figure 11. Averaged vertical profiles at 19 UTC under all sky conditions of (a) COD, (b) NO_2 photolysis rate (JNO_2), and (c) O_3 in WRF-Chem-RAP and WRF-Chem-GOES. The dashed horizontal lines indicate the average PBL height in the two simulations, which is the same. Note that the y axis shows up to 13 km in (a) but 3 km in (b) and (c).

Table 6
Performance Comparison Between WRF-Chem-HRRR and WRF-Chem-RAP

Sensitivity experiments	Cloudiness accuracy	Cloudiness POD	2-m temperature mean	PBLH mean	2-m temperature mean bias	2-m temperature RMSE
WRF-Chem-HRRR	69.5%	52.0%	28.9 °C	1.30 km	0.21 °C	2.30 °C
WRF-Chem-RAP	70.5%	54.6%	28.7 °C	1.25 km	−0.021 °C	2.25 °C

Note. The values are based on day 1 forecast averaged over CONUS for the period of 2–16 August 2016.

4.3. Effects of Satellite Corrected Clouds and Photochemistry on O₃

It is found that the clouds assimilated in RAPv3 are retained only for a few hours, even though the hourly assimilated clouds act to provide better initial conditions for pollutants in WRF-Chem-RAP. In this subsection, we assess the effects of corrected cloudiness on photochemistry and O₃ using GOES cloud retrievals in addition to hourly updated meteorology using RAPv3. The results of WRF-Chem-GOES and WRF-Chem-RAP are compared in Figure 10. Overall, surface MDA8 O₃ in WRF-Chem-GOES is higher than in WRF-Chem-RAP over CONUS and subregions. Even though the 99th percentile of MDA8 O₃ in WRF-Chem-GOES is closest to the observations over CONUS, the majority of MDA8 O₃ percentiles in WRF-Chem-GOES simulations are overpredicted compared to the observations, except for Rocky Mountain, California, and Pacific Coast where WRF-Chem-RAP had the tendency to underpredict O₃.

The primary reason for lower O₃ in WRF-Chem-RAP is due to too thick low-level clouds with cloud top height below 3 km, which results in lower JNO₂ within the boundary layer (Figure 11). In Figure 11, the peak COD in WRF-Chem-RAP (1.44) is found at 1.7 km and this value is larger by a factor of 6 than the COD from GOES retrievals at the same level (0.24). For GOES retrievals, the peak COD is 0.37 at 4.7 km. This is consistent with the results shown in Figure 2 that the model produces too many thick low-level clouds ranging from 10 to 50 and few high-level clouds as compared to GOES cloud retrievals. Due to thick low-level clouds in WRF-Chem-RAP, the boundary layer-averaged JNO₂ is lower by ~5% than that in WRF-Chem-GOES. The resulting difference in O₃ between the two simulations is ~4 ppb. In addition to overestimation of COD of low-level clouds, missing high-level clouds with CODs of 1–7.5 in the model can contribute to the lower O₃ in WRF-Chem-RAP than in WRF-Chem-GOES. As shown by Ryu et al. (2017), high-level thin clouds can increase actinic flux even at the ground level as compared to cloud-free actinic flux. This phenomenon has also been observed (e.g., Calbo et al., 2005), but to the authors' knowledge few numerical studies have reported the role of high-level thin clouds in photochemistry. More systematic study on this topic would be required in the future.

4.4. Sensitivity Simulations Using HRRR Forecasts

We conduct sensitivity simulations using HRRR forecasts that have higher spatial resolution (3 km). Using HRRR forecasts is expected to provide initial cloud conditions resolved at higher spatial scales. Small differences in cloudiness, temperature, and PBL height are found between WRF-Chem-HRRR and WRF-Chem-RAP, but in general WRF-Chem-RAP shows better agreements with observations (Table 6). Even though the day 1 O₃ forecast performance is slightly better in WRF-Chem-HRRR in terms of RMSE (10.25 ppb) than in WRF-Chem-RAP (10.32 ppb), the mean bias error is larger in WRF-Chem-HRRR (0.97 ppb) than in WRF-Chem-RAP (0.01 ppb). Overall, the better performance in meteorology in WRF-Chem-RAP leads to better performance in O₃ prediction in WRF-Chem-RAP, which supports our conclusion highlighting the important role of meteorology in O₃ prediction.

5. Summary

We evaluate and compare the surface meteorology, cloud and O₃ forecast performance over CONUS from the 12-km WRF-Chem model simulations driven by RAPv3 and GFS forecasts as initial meteorological conditions. For WRF-Chem-RAP, hourly forecasts as analogous to the RAPv3 are produced and the chemistry is cycled using the 1-hr forecast of the previous cycle.

The day 1 meteorology forecast from WRF-Chem-RAP and WRF-Chem-GFS considerably differs in near-surface temperature, dew point temperature and PBL height. The WRF-Chem-RAP corrects the cold

temperature bias that is found in WRF-Chem-GFS, and therefore performs better in predicting temperature than WRF-Chem-GFS. The higher temperature in WRF-Chem-RAP generally leads to higher PBL height due to higher soil temperature and sensible heat flux in WRF-Chem-RAP than in WRF-Chem-GFS. The evaluation of PBL height against observations at the ARM Southern Great Plain site shows a better agreement in WRF-Chem-RAP, which suggests that a better performance in temperature can lead to a better performance in PBL height. The wind predictions are found to be similar between the two simulations.

The day 1 cloud forecast (13–23 hr later after the initialization) is not substantially improved in WRF-Chem-RAP compared to WRF-Chem-GFS, implying that day 1 cloud forecasts are governed mostly by the model dynamics and physics rather than the initial meteorological conditions. The quantitative evaluation of COD and cloud top height reveals that neither WRF-Chem-RAP nor WRF-Chem-GFS is able to capture the cloud properties and distributions observed in satellite cloud retrievals. Both simulations fail to reproduce the low-level and high-level clouds with CODs of 1–7.5, but instead predict too many very thin clouds (CODs < 1) and overpredict the CODs of low-level clouds. The hourly WRF-Chem-RAP forecasts indeed show a slightly better agreement with satellite clouds, but the improvement only lasts for ~3 hr.

The model performance in forecasting next-day surface MDA8 O₃ is significantly improved over the eastern and central CONUS when RAPv3 is used by reducing positive O₃ bias found in WRF-Chem-GFS. The maximum reduction in RMSE by 5.4 ppb is obtained over the northeast in WRF-Chem-RAP compared to that in WRF-Chem-GFS. The tendency to overpredict O₃ in WRF-Chem-GFS results in a higher probability of capturing O₃ exceedances but exhibits more frequent false alarms in WRF-Chem-GFS as compared to those in WRF-Chem-RAP.

The effect of hourly updated clouds and meteorology on reducing RMSE in day 1 O₃ forecast are small under clear-sky conditions (~0.5 ppb at most), but moderate under cloudy-sky conditions (~1.5 ppb at most). Given that the initial chemistry conditions are obtained from the 1-hr forecast of the previous cycle, the hourly updated WRF-Chem simulations benefit from hourly updated meteorology because refreshing meteorology reduces errors in O₃ as well as other pollutants associated with meteorology, and this provides better initial conditions of pollutants to the next cycle.

It is found that the difference in PBL height significantly affects O₃ forecast performance between two simulations. The deeper boundary layer in WRF-Chem-RAP due to the higher temperature results in the enhanced dilution of O₃ and its precursors and also the lower chemical production of O₃.

Sensitivity simulations based on the GOES satellite cloud products show that the model's tendency to overpredict CODs of low-level clouds (and so underpredict photolysis rates) leads to lower O₃ concentration in WRF-Chem-RAP compared to WRF-Chem-GOES. These results suggest that a more skillful representation of low-level clouds is required to improve O₃ forecast. The lack of high-level thin clouds in the model could also affect lower O₃ concentrations in WRF-Chem-RAP compared to WRF-Chem-GOES.

Acknowledgments

This study is supported by NASA-ROSES Grant NNX15AE38G. The National Center for Atmospheric Research is sponsored by the National Science Foundation. We thank Chris Snyder, Rajesh Kumar, and three anonymous reviewers for their helpful comments and suggestions. We thank Mary Barth for a constructive discussion of model design associated with clouds. We would like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. The surface meteorology data are available at this site (<https://rda.ucar.edu/datasets/ds472.0/>). The boundary layer height estimates and temperature data at ARM SGP site are available at this site (<https://www.arm.gov/capabilities/observatories/sgp>). The EPA surface ozone observation data are available at this site (https://aqs.epa.gov/aqsweb/airdata/download_files.html). The satellite cloud retrievals were downloaded from this site (<https://search.earthdata.nasa.gov>). The WRF-Chem simulation outputs are available here (10.5281/zenodo.3461300).

References

- Bayler, G. M., Aune, R. M., & Raymond, W. H. (2000). NWP cloud initialization using GOES sounder data and improved modeling of nonprecipitating clouds. *Monthly Weather Review*, 128(11), 3911–3920. [https://doi.org/10.1175/1520-0493\(2001\)129<3911:nciugs>2.0.co;2](https://doi.org/10.1175/1520-0493(2001)129<3911:nciugs>2.0.co;2)
- Bei, N., de Foy, B., Lei, W., Zavala, M., & Molina, L. T. (2008). Using 3DVAR data assimilation system to improve ozone simulations in the Mexico City basin. *Atmospheric Chemistry and Physics*, 8(24), 7353–7366. <https://doi.org/10.5194/acp-8-7353-2008>
- Benedict, K. B., Zhou, Y., Sive, B. C., Prenni, A. J., Gebhart, K. A., Fischer, E. V., et al. (2019). Volatile organic compounds and ozone in Rocky Mountain National Park during FRAPPE. *Atmospheric Chemistry and Physics*, 19(1), 499–521. <https://doi.org/10.5194/acp-19-499-2019>
- Benjamin, S. G., Grell, G. A., Brown, J. M., Smirnova, T. G., & Bleck, R. (2004). Mesoscale weather prediction with the RUC hybrid isentropic–terrain-following coordinate model. *Monthly Weather Review*, 132, 473–494. [https://doi.org/10.1175/1520-0493\(2004\)132<0473:mwpwtr>2.0.co;2](https://doi.org/10.1175/1520-0493(2004)132<0473:mwpwtr>2.0.co;2)
- Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., et al. (2016). A North American hourly assimilation and model forecast cycle: The rapid refresh. *Monthly Weather Review*, 144(4), 1669–1694. <https://doi.org/10.1175/mwr-d-15-0242.1>
- Briegleb, B. P. (1992). Delta-Eddington approximation for solar radiation in the NCAR community climate model. *Journal of Geophysical Research – Atmospheres*, 97(D7), 7603–7612. <https://doi.org/10.1029/92JD00291>
- Bytheway, J. L., & Kummerow, C. D. (2015). Toward an object-based assessment of high-resolution forecasts of long-lived convective precipitation in the central U.S. *Journal of Advances in Modeling Earth Systems*, 7, 1248–1264. <https://doi.org/10.1002/2015MS000497>
- Calbo, J., Pages, D., & Gonzalez, J.-A. (2005). Empirical studies of cloud effects on UV radiation: A review. *Reviews of Geophysics*, 43 RG 2002. <https://doi.org/10.1029/2004rg000155>

- Chang, J. S., Brost, R. A., Isaksen, I. S. A., Madronich, S., Middleton, P., Stockwell, W. R., & Walcek, C. J. (1987). A three-dimensional Eulerian acid deposition model: Physical concepts and formulation. *Journal of Geophysical Research – Atmospheres*, 92(D12), 14,681–14,700. <https://doi.org/10.1029/JD092iD12p14681>
- Cheadle, L. C., Oltmans, S. J., Petron, G., Schnell, R. C., Mattson, E. J., Herndon, S. C., et al. (2017). Surface ozone in the Colorado northern Front Range and the influence of oil and gas development during FRAPPE/DISCOVER-AQ in summer 2014. *Elementa: Science of the Anthropocene*, 5, 61. <https://doi.org/10.1525/elementa.254>
- Chen, Y., Zhang, R., Meng, D., Min, J., & Zhang, L. (2016). Variational assimilation of satellite cloud water/ice path and microphysics scheme sensitivity to the assimilation of a rainfall case. *Advances in Atmospheric Sciences*, 33(10), 1158–1170. <https://doi.org/10.1007/s00376-016-6004-3>
- Eder, B., Kang, D., Mathur, R., Pleim, J., Yu, S., Otte, T., & Pouliot, G. (2009). A performance evaluation of the National Air Quality Forecast Capability for the summer of 2007. *Atmospheric Environment*, 43(14), 2312–2320. <https://doi.org/10.1016/j.atmosenv.2009.01.033>
- Eder, B., Kang, D., Mathur, R., Yu, S., & Schere, K. (2006). An operational evaluation of the Eta-CMAQ air quality forecast model. *Atmospheric Environment*, 40(26), 4894–4905. <https://doi.org/10.1016/j.atmosenv.2005.12.062>
- Foley, K. M., Hogrefe, C., Pouliot, G., Possiel, N., Roselle, S. J., Simon, H., & Timin, B. (2015). Dynamic evaluation of CMAQ part I: Separating the effects of changing emissions and changing meteorology on ozone levels between 2002 and 2005 in the eastern US. *Atmospheric Environment*, 103, 247–255. <https://doi.org/10.1016/j.atmosenv.2014.12.038>
- Gilliam, R. C., Hogrefe, C., Godowitch, J. M., Napelenok, S., Mathur, R., & Rao, S. T. (2015). Impact of inherent meteorology uncertainty on air quality model predictions. *Journal of Geophysical Research: Atmospheres*, 120, 12,259–12,280. <https://doi.org/10.1002/2015jd023674>
- Grell, G. A., & Freitas, S. R. (2014). A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry and Physics*, 14, 5233–5250. <https://doi.org/10.5194/acp-14-5233-2014>
- Hahn, C. J., Rossow, W. B., & Warren, S. G. (2001). ISCCP cloud properties associated with standard cloud types identified in individual surface observations. *Journal of Climate*, 14(1), 11–28. [https://doi.org/10.1175/1520-0442\(2001\)014<0011:ICPAWS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<0011:ICPAWS>2.0.CO;2)
- Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research – Atmospheres*, 113, D13103. <https://doi.org/10.1029/2008JD009944>
- Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A., Balzarini, A., et al. (2015). Evaluation of operational online-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part I: Ozone. *Atmospheric Environment*, 115, 404–420. <https://doi.org/10.1016/j.atmosenv.2014.09.042>
- Jones, T. A., Stensrud, D. J., Minnis, P., & Palikonda, R. (2013). Evaluation of a forward operator to assimilate cloud water path into WRF-DART. *Monthly Weather Review*, 141(7), 2272–2289. <https://doi.org/10.1175/MWR-D-12-00238.1>
- Knote, C., Hodzic, A., Jimenez, J. L., Volkamer, R., Orlando, J. J., Baidar, S., et al. (2014). Simulation of semi-explicit mechanisms of SOA formation from glyoxal in aerosol in a 3-D model. *Atmospheric Chemistry and Physics*, 14, 6213–6239. <https://doi.org/10.5194/acp-14-6213-2014>
- Lawson, J. R., Kain, J. S., Yussouf, N., Dowell, D. C., Wheatley, D. M., Knopfmeier, K. H., & Jones, T. A. (2018). Advancing from convection-allowing NWP to warn-on-forecast: Evidence of progress. *Weather and Forecasting*, 33(2), 599–607. <https://doi.org/10.1175/WAF-D-17-0145.1>
- Liu, S., & Liang, X.-Z. (2010). Observed diurnal cycle climatology of planetary boundary layer height. *Journal of Climate*, 23(21), 5790–5809. <https://doi.org/10.1175/2010jcli3552.1>
- Martinet, P., Fourrié, N., Bouteloup, Y., Bazile, E., & Rabier, F. (2014). Toward the improvement of short-range forecasts by the analysis of cloud variables from IASI radiances. *Atmospheric Science Letters*, 15(4), 342–347. <https://doi.org/10.1002/asl2.510>
- Minnis, P., Sun-Mack, S., Young, D. F., Heck, P. W., Garber, D. P., Chen, Y., et al. (2011). CERES Edition-2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS Data Part I: Algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11), 4374–4400. <https://doi.org/10.1109/tgrs.2011.2144601>
- Molod, A., Salmun, H., & Dempsey, M. (2015). Estimating planetary boundary layer heights from NOAA profiler network wind profiler data. *Journal of Atmospheric and Oceanic Technology*, 32(9), 1545–1561. <https://doi.org/10.1175/jtech-d-14-00155.1>
- Nakanishi, M., & Niino, H. (2006). An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Boundary-Layer Meteorology*, 119, 397–407. <https://doi.org/10.1007/s10546-005-9030-8>
- Nakanishi, M., & Niino, H. (2009). Development of an improved turbulence closure model for the atmospheric boundary layer. *Journal of the Meteorological Society of Japan*, 87, 895–912. <https://doi.org/10.2151/jmsj.87.895>
- Oltmans, S. J., Cheadle, L. C., Johnson, B. J., Schnell, R. C., Helmig, D., Thompson, A. M., et al. (2019). Boundary layer ozone in the Northern Colorado front range in July–August 2014 during FRAPPE and DISCOVER-AQ from vertical profile measurements. *Elementa: Science of the Anthropocene*, 7(1), 6. <https://doi.org/10.1525/elementa.345>
- Ryan, W. F. (2016). The air quality forecast rate: Recent changes and future challenges. *Journal of the Air & Waste Management Association*, 66(6), 576–596. <https://doi.org/10.1080/10962247.2016.1151469>
- Ryu, Y., Hodzic, A., Descombes, G., Hall, S., Minnis, P., Spangenberg, D., et al. (2017). Improved modeling of cloudy-sky actinic flux using satellite cloud retrievals. *Geophysical Research Letters*, 44, 1592–1600. <https://doi.org/10.1002/2016gl071892>
- Ryu, Y.-H., Hodzic, A., Barre, J., Descombes, G., & Minnis, P. (2018). Quantifying errors in surface ozone predictions associated with clouds over the CONUS: A WRF-Chem modeling study using satellite cloud retrievals. *Atmospheric Chemistry and Physics*, 18(10), 7509–7525. <https://doi.org/10.5194/acp-18-7509-2018>
- Saide, P. E., Mena-Carrasco, M., Tolvet, S., Hernandez, P., & Carmichael, G. R. (2016). Air quality forecasting for winter-time PM 2.5 episodes occurring in multiple cities in central and southern Chile: PM 2.5 FORECASTS FOR SOUTH-CENTRAL CHILE. *Journal of Geophysical Research: Atmospheres*, 121, 558–575. <https://doi.org/10.1002/2015jd023949>
- Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review*, 136, 5095–5115. <https://doi.org/10.1175/2008mwr2387.1>
- van der Veen, S. H. (2013). Improving NWP model cloud forecasts using Meteosat second-generation imagery. *Monthly Weather Review*, 141, 1545–1557. <https://doi.org/10.1175/mwr-d-12-00021.1>
- Wilczak, J. M., Djalalova, I., McKeen, S., Bianco, L., Bao, J.-W., Grell, G., et al. (2009). Analysis of regional meteorology and surface ozone during the TexAQS II field program and an evaluation of the NMM-CMAQ and WRF-Chem air quality models. *Journal of Geophysical Research – Atmospheres*, 114, D00F14. <https://doi.org/10.1029/2008jd011675>
- Wood, R. (2012). Stratocumulus clouds. *Monthly Weather Review*, 140(8), 2373–2423. <https://doi.org/10.1175/MWR-D-11-00121.1>

- Yucel, I., James Shuttleworth, W., Gao, X., & Sorooshian, S. (2003). Short-term performance of MM5 with cloud-cover assimilation from satellite observations. *Monthly Weather Review*, 131(8), 1797–1810. <https://doi.org/10.1175/2565.1>
- Yucel, I., Shuttleworth, W. J., Pinker, R. T., Lu, L., & Sorooshian, S. (2002). Impact of ingesting satellite-derived cloud cover into the regional atmospheric modeling system. *Monthly Weather Review*, 130(3), 610–628. [https://doi.org/10.1175/1520-0493\(2002\)130<0610:ioisdc>2.0.co;2](https://doi.org/10.1175/1520-0493(2002)130<0610:ioisdc>2.0.co;2)
- Zhang, F., Bei, N., Nielsen-Gammon, J. W., Li, G., Zhang, R., Stuart, A., & Aksoy, A. (2007). Impacts of meteorological uncertainties on ozone pollution predictability estimated through meteorological and photochemical ensemble forecasts. *Journal of Geophysical Research – Atmospheres*, 112(D4). <https://doi.org/10.1029/2006jd007429>
- Zhou, G., Xu, J., Xie, Y., Chang, L., Gao, W., Gu, Y., & Zhou, J. (2017). Numerical air quality forecasting over eastern China: An operational application of WRF-Chem. *Atmospheric Environment*, 153, 94–108. <https://doi.org/10.1016/j.atmosenv.2017.01.020>